

PharmaGist: a webserver for ligand-based pharmacophore detection

Dina Schneidman-Duhovny^{1,*}, Oranit Dror¹, Yuval Inbar¹, Ruth Nussinov^{2,3} and Haim J. Wolfson¹

¹School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, ²Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel and ³Basic Research Program, SAIC-Frederick, Center for Cancer Research, Nanobiology Program, NCI-Frederick, Bldg 469, Rm 151, Frederick, MD 21702, USA

Received February 4, 2008; Revised March 24, 2008; Accepted March 30, 2008

ABSTRACT

Predicting molecular interactions is a major goal in rational drug design. Pharmacophore, which is the spatial arrangement of features that is essential for a molecule to interact with a specific target receptor, is an important model for achieving this goal. We present a freely available web server, named *PharmaGist*, for pharmacophore detection. The employed method is ligand based. Namely, it does not require the structure of the target receptor. Instead, the input is a set of structures of drug-like molecules that are known to bind to the receptor. The output consists of candidate pharmacophores that are computed by multiple flexible alignment of the input ligands. The method handles the flexibility of the input ligands explicitly and in deterministic manner within the alignment process. PharmaGist is also highly efficient, where a typical run with up to 32 drug-like molecules takes seconds to a few minutes on a standard PC. Another important characteristic is the capability of detecting pharmacophores shared by different subsets of input molecules. This capability is a key advantage when the ligands belong to different binding modes or when the input contains outliers. The webserver has a user-friendly interface available at <http://bioinfo3d.cs.tau.ac.il/PharmaGist>.

INTRODUCTION

A *pharmacophore* is the spatial arrangement of features that enables a molecule to interact with a target receptor

in a specific binding mode. Once identified, a pharmacophore can serve as a powerful model in versatile applications for rational drug design, such as virtual screening, *de novo* design, lead optimization and ADME/Tox studies (1). In addition, pharmacophore models can be useful for Chemogenomics studies (2,3). This is a new field for systematically studying the effect of a large number of drug-like molecules on a biological network of macromolecular targets. Pharmacophores of target macromolecules in a network can be used to explain and predict the biological effect of new drug-like ligands. Indeed, a large-scale identification of pharmacophores for many biological macromolecules is becoming more and more feasible due to the expanding number of databases of drug-like ligands with known interacting receptors.

Many computational methods for pharmacophore identification have been introduced (1,4). Some methods, which are called the *direct methods*, use both ligand and receptor information. However, in most cases the structure of the target receptor is unknown and the only available information is a set of ligands that have been observed to interact with the receptor. In such cases, only the second type of methods, the *indirect methods*, are applicable, since they are based on ligand information alone.

Given a set of drug-like molecules that are known to interact with a specific target receptor, the aim of the indirect methods is to find the largest (or highest scoring) 3D pattern of features responsible for binding that is shared by all or most of the input ligands. The task is challenging due to the hardness of the problem with respect to both the number of input molecules and the flexibility of drug-like molecules. Specifically, aligning drug-like molecules is NP hard even in the case of three rigid molecules (5,6) and in the case of a pair of molecules, where one is rigid and the other one is flexible (7).

*To whom correspondence should be addressed. Tel/Fax: +972 3 640 6476; Email: duhovka@tau.ac.il

The publisher or recipient acknowledges right of the U.S. Government to retain a nonexclusive, royalty-free license in and to any copyright covering the article.

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors

Existing indirect methods for pharmacophore detection mainly differ in their approach for addressing the flexibility of a drug-like molecule. A drug-like molecule may possess several rotatable bonds and as a result it may have many potentially feasible conformations. The specific conformations that the input ligands adopt when interacting with the receptor are unknown. Therefore, all their feasible conformations have to be considered. Most indirect methods perform the conformational search as a separate initial stage. Specifically, a discrete set of conformations is generated for each input ligand with the aim of sampling all its feasible conformations. Examples for such methods include RAPID (8), MPHIL (9) and the commercial programs DISCO (10), Phase (11) of Schrodinger and HipHop and HypoGen, which are part of the Catalyst software package of Accelrys (12–16). The main limitation of this approach is that covering the whole conformational space might require an extremely large number of conformations, especially for molecules with many rotatable bonds. An alternative approach is to incorporate the conformational search within the pattern identification process. This way the search space is not restricted to a precompiled discrete set of conformations. However, the current methods that adopt this approach are based on random search. Among them are SCAMPI (17), GAMMA (18–20), GASP (21,22) and GALAHAD (23,24) of Tripos and the MOGA-based method described in ref. (25).

Another approach for treating the flexibility of the ligands within the pattern detection process is employed by the MTree method of BioSolveIT, which aligns feature-trees derived from the 2D structures of the ligands (26).

Herein, we present PharmaGist, the first webserver for elucidating 3D pharmacophores from a set of drug-like molecules that are known to bind to a target receptor. The method efficiently searches for possible pharmacophores and reports the highest-scoring ones (7). The candidate pharmacophores are detected by multiple flexible alignment of the input ligands, where the flexibility of the ligands is treated explicitly and in a deterministic manner in the alignment process. Another key advantage of the method is the ability to detect pharmacophores common to subsets of input ligands, a characteristic that makes PharmaGist tolerant to outliers and to several binding modes. The performance of PharmaGist was successfully evaluated on different test cases (7). The whole dataset consists of almost 80 crystal structures of receptor–ligand complexes taken mainly from the FlexS benchmark dataset (27). The complexes are classified into 12 different cases, where each case includes complexes of the same receptor with different ligands. For validation, we produced a reference pharmacophore for each test case. The reference pharmacophore was extracted from a 3D alignment of the bounded ligands. This alignment was derived from a superposition of the receptor in the different complexes. In all cases, the highest scoring pharmacophore candidate detected by PharmaGist was similar to the reference pharmacophore. The runtime ranges from seconds to a few minutes on a typical set of ligands. Below, we briefly describe the method followed by a comprehensive description of the webserver. The interface of the server is user friendly with only two mandatory input fields. The output provides a convenient

browsing over the suggested pharmacophores and an interactive visualization using Jmol (28).

METHOD OVERVIEW

The input is set of drug-like ligands in a 3D representation. The output is a list of candidate pharmacophores. These are 3D patterns of physico-chemical features that are shared by all or some input ligands. In addition, the output provides for each candidate pharmacophore a 3D superposition of conformations of input ligands that share it. Generally, the algorithm suggests solving the task by multiple flexible alignment of drug-like molecules. However, due to the hardness of the problem, a heuristic solution is provided in practice. In addition, the algorithm assumes that one of the input ligands is given in a conformation that is not very different from the bound state. This ligand serves as a *pivot* on which the other *target* ligands are aligned. By default, the algorithm iteratively selects each input ligand to serve as a pivot. Alternatively, the pivot can be defined by the user as the ligand with the highest affinity to the receptor or the one with the lowest number of rotatable bonds. Also, in case of a highly flexible pivot, it is possible to use a pre-computed set of conformations.

The method consists of four major stages: (i) ligand representation, (ii) pairwise alignment, (iii) multiple alignment and (iv) solution clustering and output (Figure 1). In the first stage, each input ligand is processed separately. The method detects the rotatable bonds of the ligand and divides it into rigid groups accordingly. In addition, the ligand is assigned with a set of physico-chemical features (hydrogen bond donor/acceptor, anion/cation, aromatic ring, hydrophobic group or, optionally, other features defined by the user). In the second stage, given a pivot (treated as rigid) and one target ligand (treated as flexible), pairwise alignments are computed as follows. First, for each rigid group of the target ligand, the method generates a set of transformations for superimposing the target rigid group onto the pivot. The result for each target rigid group is a set of candidate new poses on the pivot. Then, these poses are reassembled into new conformations of the target ligand aligned on the pivot. The score of a resulting pairwise alignment is a weighted sum of the matched pivot features. Two features, one from the pivot and one from a conformation of the target ligand, can be matched if they are of the same type and the distance between them is below a predefined threshold (1 Å by default). The algorithm uses default weight values for each feature type (0.3 for hydrophobicity and 1 for the rest), unless other values are supplied by the user. The output of the stage is a large number of high-scoring pairwise alignments between the pivot and the target ligand. The third stage also works with a selected pivot. Pairwise alignments between the pivot and the target ligands are combined into multiple alignments. The goal is to find significant subsets of pivot features that are matched by as many pairwise alignments for different target ligands as possible. However, maximizing the number of aligned ligands can be contradictory to

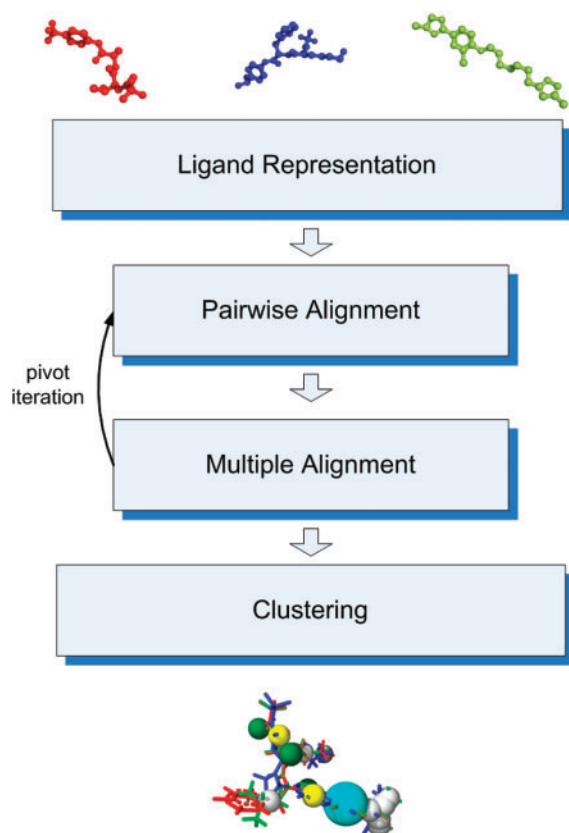


Figure 1. PharmaGist method flow.

maximizing the score of the matched features. Thus, the method produces multiple alignments for each subset size of input ligands. Due to efficiency considerations, this is achieved by enumerating all the possible subsets of pivot features and selecting the ones that can be aligned by as many ligands as possible. Subsets of pivot features with a significant score are candidate pharmacophores and will be reported to the user. By default, a candidate pharmacophore must consist of at least three spatially distinct features, but this parameter is user-defined. In the last fourth stage, candidate pharmacophores derived from different pivot iterations are clustered and the highest scoring non-redundant ones for each number of molecules are reported.

WEBSERVER

The web interface is very simple. Its input form has only two mandatory fields: a file with input ligands and an email address. After submitting the form, the pharmacophore detection algorithm starts running. When the run completes, a link to a web page with candidate pharmacophores is sent to the user by an email.

Input

The basic input form is simple with two mandatory fields: a file with drug-like molecules in Mol2 format and an email address. Currently, the user may upload up to

32 molecules, either in a single Mol2 file or in one zip file compressing multiple Mol2 files. The basic form has an additional field for setting the number of candidate pharmacophores to be reported for each subset size of input molecules (10 by default). Other parameters can be set in the 'Advanced Options' part of the form. By default, these parameters are hidden and setting them is optional. Users are encouraged to change these parameters if additional data on the specific-binding site of the target receptor are known. The advanced options include selecting a key molecule to serve as a pivot and defining the minimal number of spatially distinct features in a pharmacophore. Additionally, users can set the scoring weight assigned to each feature type. For example, if the interaction with the target receptor is mostly hydrophobic, the scoring weight assigned to hydrophobic features can be increased. Finally, the server allows users to define their own feature types with assigned scoring weights. This option is useful when it is known that a specific group is essential for an interaction with the receptor. An additional explanation on each field is available in the help page, which can be directly accessed by clicking on the name of the field.

Output

A typical run of PharmaGist takes seconds to a few minutes. Once the run completes, a link to a web page with an access to the results is sent to the user via an email. The results are stored on the server for at least a month, so the user can view the results later again using the same link.

The output of PharmaGist is demonstrated for an input with seven elastase inhibitors. Figure 2 presents the main output page. This page contains several tables. The upper table lists the input molecules with their number of atoms and assigned physico-chemical features. At the top of the table, there is link to a web page that displays the molecules with their features in a Jmol applet (28). The next tables summarize the results. Each table presents the highest scoring candidate pharmacophores shared by a specific number of molecules. The tables are sorted in descending order by the number of molecules. The candidate pharmacophores in each table are displayed in descending order by their scores. Note that at the top of the result page there is a link to a page in which all the candidate pharmacophores are sorted by their score, regardless of the number of ligands that share them. At the bottom of the result page, there is a link to a page that displays the best flexible alignment for each pair of input molecules. In the result tables, the following data are presented for each candidate pharmacophore: (i) the score, (ii) the names of the aligned molecules, (iii) the number of common features and their type distribution and (iv) a link to a page that displays the candidate pharmacophore.

Figure 3 shows an example for a page that describes a single candidate pharmacophore. Specifically, the page describes the top scoring candidate pharmacophore shared by all seven elastase inhibitors. The page supplies a 3D visualization of the pharmacophore and the multiple

Input Molecules [view details: visualization of the detected features](#)

#	Molecule	Atoms	Features	Spatial Features	Aromatic	Hydrophobic	Donors	Acceptors	Negatives	Positives
1	elastase_1.mol2	64	16	15	1	8	3	3	0	1
2	elastase_2.mol2	69	21	20	1	12	4	3	0	1
3	elastase_3.mol2	70	18	17	2	8	4	3	0	1
4	elastase_4.mol2	52	10	10	2	2	3	3	0	0
5	elastase_5.mol2	48	12	12	1	5	3	3	0	0
6	elastase_6.mol2	37	11	11	3	1	1	6	0	0
7	elastase_7.mol2	42	14	14	2	3	2	6	1	0

[Sort by score](#)

Number of Aligned Molecules: 7

Score	Jmol	Features	Spatial Features	Aromatic	Hydrophobic	Donors	Acceptors	Negatives	Positives	Molecules
19.843	j7_1.html	4	4	1	0	1	2	0	0	elastase_3.mol2 elastase_1.mol2 elastase_2.mol2 elastase_4.mol2 elastase_5.mol2 elastase_6.mol2 elastase_7.mol2
16.668	j7_2.html	5	5	0	1	1	3	0	0	elastase_7.mol2 elastase_1.mol2 elastase_2.mol2 elastase_3.mol2 elastase_4.mol2 elastase_5.mol2 elastase_6.mol2

Number of Aligned Molecules: 6

Score	Jmol	Features	Spatial Features	Aromatic	Hydrophobic	Donors	Acceptors	Negatives	Positives	Molecules
19.106	j6_1.html	5	5	1	1	1	2	0	0	elastase_7.mol2 elastase_1.mol2 elastase_2.mol2 elastase_3.mol2 elastase_4.mol2 elastase_5.mol2
19.106	j6_2.html	5	5	1	1	1	2	0	0	elastase_7.mol2 elastase_2.mol2 elastase_3.mol2 elastase_4.mol2 elastase_5.mol2 elastase_6.mol2

Number of Aligned Molecules: 5

Score	Jmol	Features	Spatial Features	Aromatic	Hydrophobic	Donors	Acceptors	Negatives	Positives	Molecules
24.820	j5_1.html	8	8	1	2	2	3	0	0	elastase_5.mol2 elastase_1.mol2 elastase_2.mol2 elastase_3.mol2 elastase_4.mol2
20.795	j5_2.html	6	6	1	1	2	2	0	0	elastase_7.mol2 elastase_2.mol2 elastase_3.mol2 elastase_4.mol2 elastase_5.mol2

Figure 2. Main output page. The figure displays parts of the main output page obtained for an input with seven elastase inhibitors. Note that the link to the page that displays the best flexible alignment for each pair of input molecules is not shown due to space limitation. This link appears at the bottom of the main output page.

Pharmacophore features: ☒

Score	Features	Spatial Features	Aromatic	Hydrophobic	Donors	Acceptors	Negatives	Positives
19.843	4	4	1	0	1	2	0	0

Molecule Name	Show Molecule:	Show Features:
elastase_3*	<input checked="" type="checkbox"/>	<input type="checkbox"/>
elastase_1	<input checked="" type="checkbox"/>	<input type="checkbox"/>
elastase_2	<input checked="" type="checkbox"/>	<input type="checkbox"/>
elastase_4	<input checked="" type="checkbox"/>	<input type="checkbox"/>
elastase_5	<input checked="" type="checkbox"/>	<input type="checkbox"/>
elastase_6	<input checked="" type="checkbox"/>	<input type="checkbox"/>
elastase_7	<input checked="" type="checkbox"/>	<input type="checkbox"/>

* pivot molecule

[Download Pharmacophore and Alignment File](#)

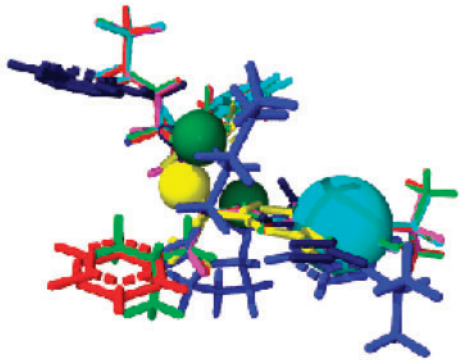


Figure 3. Pharmacophore output page. The displayed page describes the top scoring candidate pharmacophore shared by seven elastase inhibitors.

flexible alignment that it is based on. The page consists of three main parts: (i) a summary of the attributes of the pharmacophore at the top, (ii) a Jmol display on the right and (iii) a display control panel on the left. The summary part contains the score of the candidate pharmacophore, its number of features and their physico-chemical types. By default, the Jmol display shows the pharmacophore (spacefill model) and the multiple flexible alignment that yields it (wireframe model). The color coding of the features in the display is by their physico-chemical types and it is as in the summary part of the page at the top. The display control panel contains one row for each molecule in the alignment. Each molecule is depicted in a different color as it appears in the Jmol display. The pivot is marked with an asterisk. Two checkboxes are provided for each molecule in the control panel. The first checkbox allows the user to show/hide the controlled molecule in the display and the second checkbox controls the display of the features of the molecule. In addition, the user can show/hide the pharmacophore in the display using the checkbox at the top of the summary part. At the bottom of the page, there is a link for downloading a Mol2 file with the candidate pharmacophore and the corresponding multiple flexible alignment.

SUMMARY

We have presented the first webserver for detecting 3D pharmacophores shared by known active ligands in the absence of structural information on the target receptor. The deterministic and efficient algorithm behind the server allows a fast and reliable detection of pharmacophores with explicit consideration of the flexibility of the ligands. A key characteristic of the server is the ability to detect outlier molecules as well as to find pharmacophores common only to subsets of input ligands. The web interface is simple and easy to use. On the one hand, the algorithm is fully automated and the user is only required to upload the input ligands. On the other hand, the server provides advanced customization options, such as defining new features or setting the scoring weight of the features. The pharmacophores found by the server can be further used for generating new leads or for optimizing existing ones in a drug design process.

ACKNOWLEDGEMENTS

We want to thank D. Fishlovitch, A. Oron and H. Senderowitz for useful advises. The research of O.D. and Y.I. has been supported by the Eshkol Fellowship funded by the Israeli Ministry of Science. The research of H.J.W. has been supported in part by the Israel Science Foundation (grant no. 281/05) and by the Hermann Minkowski-Minerva Center for Geometry at TAU. The research of H.J.W. and R.N. has been by the NIAID, NIH (grant No. 1UC1AI067231), and by the Binational US-Israel Science Foundation (BSF). This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number N01-CO-12400.

The content of this publication does not necessarily reflect the view of the policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organization imply endorsement by the U.S. Government. This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Funding to pay the Open Access Publication charges for this article was provided by NCI contract N01-CO-12400.

Conflict of interest statement. None declared.

REFERENCES

- Güner, O.F. (ed.) (2000) *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line, La Jolla, CA.
- Rognan, D. (2007) Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.*, **152**, 38–52.
- Klabunde, T. (2007) Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.*, **152**, 5–7.
- Dror, O., Shulman-Peleg, A., Nussinov, R. and Wolfson, H. (2006) Predicting molecular interactions in silico: I. an updated guide to pharmacophore identification and its applications to drug design. *Front. Med. Chem.*, **3**, 551–584.
- Akutsu, T. and Halldorsson, M.M. (2000) On the approximation of largest common subtrees and largest common point sets. *Theor. Comp. Sci.*, **233**, 33–50.
- Shatsky, M., Shulman-Peleg, A., Nussinov, R. and Wolfson, H. (2006) The multiple common point set problem and its application to molecule binding pattern detection. *J. Comp. Biol.*, **13**, 407–442.
- Inbar, Y., Schneidman-Duhovny, D., Dror, O., Nussinov, R. and Wolfson, H.J. (2007) Deterministic pharmacophore detection via multiple exible alignment of drug-like molecules. In Speed, T.P. and Huang, H., (eds), *Research in Computational Molecular Biology (RECOMB), 11th Annual International Conference*. Springer, Oakland, CA, USA, Vol. 4453, pp. 412–429.
- Finn, P.W., Kavraki, L.E., Latombe, J.-C., Motwani, R., Shelton, C., Venkatasubramanian, S. and Yao, A. (1998) RAPID: randomized pharmacophore identification for drug design. *Comp. Geom. Theor. Appl.*, **10**, 263–272.
- Holliday, J. and Willet, P. (1997) Using a genetic algorithm to identify common structural features in sets of ligands. *J. Mol. Graph. Model.*, **15**, 203–253.
- Martin, Y., Bures, M., Dahaner, E., DeLazzer, J., Lico, I. and Pavlik, P. (1993) A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput. Aided Mol. Des.*, **7**, 83–102.
- Dixon, S., Smondyrev, A., Knoll, E., Rao, S., Shaw, D. and Friesner, R. (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.*, **20**, 647–671.
- Barnum, D., Greene, J., Smellie, A. and Sprague, P. (1996) Identification of common functional configurations among molecules. *J. Chem. Inform. Comp. Sci.*, **36**, 563–571.
- Clement, O.A. and Mehl, A.T. (2000) HipHop: pharmacophores based on multiple common-feature alignments. In Güner, O.F. (ed.), *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line, La Jolla, CA, pp. 69–84.
- Li, H., Sutter, J. and Hoffmann, R. (2000) HypGen: an automated system for generating 3D predictive pharmacophore models, In Güner, O.F. (ed.), *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line, La Jolla, CA, pp. 171–189.
- Kurogi, Y. and Güner, O.F. (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr. Med. Chem.*, **8**, 1035–1055.
- Güner, O.F., Clement, O. and Kurogi, Y. (2004) Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr. Med. Chem.*, **11**, 2991–3005.

17. Chen,X., Rusinko III,A., Tropsha,A. and Young,S. (1999) Automated pharmacophore identification for large chemical data sets. *J. Chem. Inf. Comput. Sci.*, **39**, 887–896.
18. Handschuh,S., Wagener,M. and Gasteiger,J. (1997) Superimposition of three-dimensional chemical structures allowing for conformational flexibility by a hybrid method. *J. Chem. Inf. Comput. Sci.*, **38**, 220–232.
19. Handschuh,S., Wagener,M. and Gasteiger,J. (2000) The search for the spatial and electronic requirements of a drug. *J. Mol. Model.*, **6**, 358–378.
20. Handshuh,S. and Gasteiger,J. (2000) Pharmacophores derived from the 3D substructure perception, In Güner,O.F. (ed.), *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line La Jolla, CA, pp. 429–454.
21. Jones,G., Willett,P. and Glen,R. (1995) A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput. Aided Mol. Des.*, **9**, 532–549.
22. Jones,G., Willett,P. and Glen,R. (2000) GASP: genetic algorithm superposition program, In Güner,O.F. (ed.), *Pharmacophore Perception, Development, and Use in Drug Design*. International University Line, La Jolla, CA, pp. 85–106.
23. Richmond,N., Abrams,C., Wolohan,P., Abrahamian,E., Willett,P. and Clark,R. (2006) GALAHAD: 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D. *J. Comput. Aided Mol. Des.*, **20**, 567–587.
24. Shepphird,J.K. and Clark,R.D. (2006) A marriage made in torsional space: using GALAHAD models to drive pharmacophore multiplet searches. *J. Comput. Aided Mol. Des.*, **20**, 763–771.
25. Cottrell,S.J., Gillet,V.J., Taylor,R. and Wilton,D.J. (2004) Generation of multiple pharmacophore hypotheses using multi-objective optimisation techniques. *J. Comput. Aided Mol. Des.*, **18**, 665–682.
26. Hessler,G., Zimmermann,M., Matter,H., Evers,A., Naumann,T., Lengauer,T. and Rarey,M. (2005) Multiple-ligand-based virtual screening: methods and applications of the MTree approach. *J. Med. Chem.*, **48**, 6575–6584.
27. Lemmen,C., Lengauer,T. and Klebe,G. (1998) FlexS: a method for fast flexible ligand superposition. *J. Med. Chem.*, **41**, 4502–4520.
28. Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org> (1 January 2008, date last accessed).